An A Contrario Model for Matching Interest Points under Geometric and Photometric Constraints

Frédéric Sur, Nicolas Noury, Marie-Odile Berger

SIAM J. IMAGING SCIENCES
Vol. 6, No. 4, pp. 1956–1978

An A Contrario Model for Matching Interest Points under Geometric and Photometric Constraints*

Frédéric Sur†, Nicolas Noury†, and Marie-Odile Berger†

Abstract. Finding point correspondences between two views is generally based on the matching of local photometric descriptors. A subsequent geometric constraint ensures that the set of matching points is consistent with a realistic camera motion. Starting from a paper by Moisan and Stival, we propose an a contrario model for matching interest points based on descriptor similarity and geometric constraints. The resulting algorithm has adaptive matching thresholds and is able to detect point correspondences whose associated descriptors are not the first nearest neighbor. We also discuss the specific difficulties raised by images containing repeated patterns which are likely to introduce correspondences beyond the nearest neighbor.

Key words. point correspondence problem, a contrario model, generalized RANSAC, repeated patterns

AMS subject classifications. 62H35, 68T45

DOI. 10.1137/120871766

1. Introduction.

Matching interest points between two (or more) views is one of the keystones of many computer vision applications. "Matching" means associating interest points across several images that correspond to the same actual three-dimensional (3D) point. This is often achieved by taking into account local descriptors, i.e., an encoding of the gray values from a region around each interest point, achieved in the following manner:

1. In both views, extract interest points together with a descriptor.
2. Match them using some (dis-)similarity measure over the descriptors.
3. Keep the most consistent set of points with respect to the geometry imposed by a realistic camera motion.

The extraction of interest points and local photometric descriptors in step 1 has been the subject of a vast literature (see the reviews [1, 15, 16, 18]). The regions from which the descriptors are extracted should be covariant with some geometric transformations (and hence the descriptors should be invariant to these transformations), so that matching is robust to camera viewpoint changes. Most of the time, descriptors are invariant to affine transformations, or to similarity transformations as in Lowe's scale invariant feature transform (SIFT) [14]. In step 2, a popular way [14] to define a set of correspondences is to keep the descriptor's nearest neighbor and to estimate the reliability of a tentative correspondence by the ratio of the Euclidean distances to the nearest and the second nearest neighbor, which should be below some threshold. Step 3 is often achieved by random sample consensus (RANSAC) [10] or one of the numerous methods derived from it such as generalized RANSAC [29]. Note

*Received by the editors March 29, 2012; accepted for publication (in revised form) July 10, 2013; published electronically October 15, 2013.
http://www.siam.org/journals/siims/6-4/87176.html

†CNRS, LORIA, Université de Lorraine, 54506 Vandœuvre-lès-Nancy cedex, France (sur@loria.fr, noury@loria.fr, berger@loria.fr).

that when running RANSAC-like algorithms and especially generalized RANSAC, the user needs to tune one or more parameters.

However, considering descriptor matching and geometric enforcement as independent steps may be problematic. Of course, RANSAC keeps only a subset of the correspondences that are consistent with respect to descriptor similarity and cannot add new correspondences. Specific problems also occur with images where repeated patterns are present, like windows on a facade or a repetitive texture. In this case, invariant regions around interest points are likely to produce numerous similar descriptors, confusing nearest neighbor matching. The output of a subsequent RANSAC can then be a set of correspondences which is actually shifted when confronted with the ground truth. Even if the camera motion is properly estimated, it is possible to obtain many false correspondences consisting of pairs of points with similar descriptors but falling near the epipolar lines "by chance." Such situations are illustrated in [25].

One way to enhance the distinguishability of the descriptors even with repeated structures would be to increase the size of the underlying regions as in [28] or in the sequential correspondence verification with cosegmentation [5]. Incorporating scale information in the matching process as in [27] also enhances the distinguishability. Disambiguating correspondences with repeated patterns can also be achieved by taking account of the local organization of the interest points. In [6] SIFT descriptors are associated with region context descriptors encoding the relative position of nearby interest points. The "reinforcement matching" directly takes into account geometric information from the "region context." Hence, to some extent, it should show some robustness to repeated patterns, although no evidence is given in [6]. In the same spirit, a matching process proposed in [2] is based on the observation that the relative position of the interest points is preserved in both views, provided that the viewpoint change is limited. This algorithm was later used in [13] to make Morel and Yu's affine-SIFT (ASIFT) [19] robust to repeated patterns. However, the geometric constraints are not explicitly implemented.

Moreover, when the nearest neighbor matching is spoilt by false correspondences, it is useful to be able to obtain correspondences beyond the nearest neighbors. Generalized RANSAC [29] aims at achieving this goal but needs the tuning of several parameters. The work of [24] makes use of geometric and appearance priors to guide a homography search. Although it requires tuning several parameters, it gives good results with repeated patterns and even with quite strong viewpoint changes. In the paper introducing guided-MLESAC [26], the possibility of discarding the separated photometric matching step by incorporating photometric information (cross-correlation in [29]) in the prior is mentioned. Theoretically, this should improve the algorithm's ability to deal with repeated patterns.

The proposed contribution is in line with this approach and aims at detecting correspondences beyond the nearest neighbor, based on a geometric prior. The contribution of this paper is to design an a contrario model for matching interest points based on a metric, namely the number of false alarms (NFA), which simultaneously integrates descriptor similarity and geometric constraints. This model was first briefly suggested in a paper by Moisan and Stival [17] (under the term *colored rigidity*), but it was not investigated further. While the a contrario model of [17] deals only with the epipolar constraint, we extend it to the planar homography case. We also explain how to design an appropriate distribution function over the descriptor

dissimilarity measure for this a contrario model, based on [20, 22]. An algorithm is built to find sets of correspondences with a low NFA. It is able to detect correspondences beyond the first nearest neighbor.

This paper is organized as follows. Section 2 discusses the improved statistical a contrario model. An algorithm based on this model is designed in section 3 to replace the above-mentioned steps 2 and 3. While the matching thresholds are automatically derived from the NFA, two modeling parameters are needed. An experiment on synthetic data shows how to set them and suggests that they have a limited effect on the algorithm's output. Section 4 describes the experimental assessment and provides a proof of concept focusing on images containing repeated patterns. The a contrario algorithm represents an improvement on RANSAC based on nearest neighbor matching. It performs as well as generalized RANSAC [29], which needs parameter tuning.

2. An improved a contrario model for point correspondences under epipolar and photometric constraints.

The proposed a contrario model is based on Moisan and Stival's a contrario RANSAC [17] and on Rabin, Delon, and Gousseau's a contrario SIFT matching using the earth mover's distance [22]. Moisan and Stival [17] focus on geometric constraints and assume that correspondences between interest points are given by some prior step. Rabin, Delon, and Gousseau [22] deal only with descriptor matching and do not take into account geometric constraints. Our contribution is to merge these two approaches in an a contrario model which simultaneously takes into account the epipolar and photometric constraints.

Let us begin with some notation. We assume that two views (images $\mathcal{I}_1$ and $\mathcal{I}_2$) from the same scene are given. For each image, some algorithm (for example, SIFT) gives a set of interest points, with an invariant descriptor. Let us denote by $(x_i, D(x_i))_{1\leq i\leq N_1}$ (resp., $(y_j, D(y_j))_{1\leq j\leq N_2}$) the N_1 (resp., N_2) interest points with descriptors from $\mathcal{I}_1$ (resp., $\mathcal{I}_2$) such that x_i (resp., y_j) is the coordinate vector of an interest point and $D(x_i)$ (resp., $D(y_j)$) is the corresponding local descriptor. We denote by x_i the interest point itself, i.e., its pixel coordinates or its homogeneous coordinates in the projective plane.

Assuming a pinhole camera model, if x_i and y_j are the projections in $\mathcal{I}_1$ and $\mathcal{I}_2$ of the same 3D point, then y_j lies on the epipolar line associated with x_i. This line is represented by Fx_i, where F is the fundamental matrix from $\mathcal{I}_1$ to $\mathcal{I}_2$. Conversely, x_i has to lie on the epipolar line $F^T y_j$ since the fundamental matrix from $\mathcal{I}_2$ to $\mathcal{I}_1$ is the transpose matrix F^T. However, if the camera has just been rotated around its optical center, or if interest points lie on a common plane, then the fundamental matrix is not defined. In this case, there is a two-dimensional projective transformation (a homography) H such that $y_j = H(x_i)$ and $x_i = H^{-1}(y_j)$.

The problem of interest is therefore to find a subset $\mathcal{S}$ of $\{1,\dots,N_1\}\times\{1,\dots,N_2\}$ and a fundamental matrix F or a homography H from $\mathcal{I}_1$ to $\mathcal{I}_2$ such that the following hold:

1. The distance between corresponding descriptors is below some threshold δ_D, ensuring that the local image regions are alike:

$$\forall (i,j)\in\mathcal{S},\quad d_D(D(x_i),D(y_j))\leq\delta_D. \tag{2.1}$$

2. The distance between a point and the epipolar line associated with the corresponding point is below some other threshold δ_G (and vice versa), ensuring that the epipolar

constraint is satisfied:

$$\forall (i,j) \in \mathcal{S}, \quad d_G(x_i, y_j, F) := \max\{d_G(y_j, Fx_i), d_G(x_i, F^T y_j)\} \leq \delta_G. \tag{2.2}$$

Alternatively, for the homography constraint,

$$\forall (i,j) \in \mathcal{S}, \quad d_G(x_i, y_j, H) := \max\{d_G(y_j, H(x_i)), d_G(x_i, H^{-1}(y_j))\} \leq \delta_G. \tag{2.3}$$

The proposed statistical framework automatically balances geometry and photometry, and also automatically gives both thresholds δ_D and δ_G relative to a set $\mathcal{S}$. Before specifying distances d_D and d_G, we explain the model.

2.1. The a contrario model. Since the seminal paper by Desolneux, Moisan, and Morel [8], a contrario models have been the subject of a large literature. The books [4] and [9] and the references therein give a comprehensive account of their use in many different computer vision problems. See also [7] for a simple introduction. The idea behind a contrario models is that independent, structureless random features can produce structured groups only with a very small probability. Groups of features are said to be *meaningful* if their probability is very low under the hypothesis $\mathcal{H}_0$ that the features are independent. The independence assumption makes the probability computation tractable, since joint laws are simply products of marginal laws which can be reliably estimated with a limited number of empirical observations. In the statistical hypothesis testing framework, this probability is called a p-value: if it is low (typically below 5%), then it is likely that the group of interest does not satisfy the independence assumption $\mathcal{H}_0$. There must be a better explanation than independence for this group, and this explanation should emphasize some common causality.

Here pairs of features form a meaningful group because the descriptors are similar, and the motion of the interest points is consistent with the motion of the camera between the two views. Let us assume that a set $\mathcal{S}$ of correspondences is given, as well as a transformation A (fundamental matrix or homography) and two thresholds δ_D and δ_G as in (2.1) and (2.2)–(2.3). The probability to estimate is

$$p(\mathcal{S}, A, \delta_G, \delta_D) := \Pr\big(\forall (i,j) \in \mathcal{S}, d_G(x_i, y_j, A) \leq \delta_G \text{ and } d_D(D(x_i), D(y_j)) \leq \delta_D \mid \mathcal{H}_0\big). \tag{2.4}$$

Let us also assume that the transformation A is estimated from a minimal subset of $\mathcal{S}$ as in the RANSAC paradigm. This means in the $A = F$ case that a subset s from $\mathcal{S}$ made of $m = 7$ correspondences is used to estimate F [11]. In the $A = H$ case, $m = 4$ points are needed in s. In what follows, $S \backslash s$ is the set of correspondences in S that are not in s.

Definition 2.1. *Considering $(x_i, D(x_i))$ and $(y_j, D(y_j))$ as random variables, we define hypothesis $\mathcal{H}_0$ as follows:*

1. *$(d_D(D(x_i), D(y_j)))_{(i,j)\in\mathcal{S}}$ and $(d_G(x_i, y_j, A))_{(i,j)\in\mathcal{S}\backslash s}$ are mutually independent random variables.*
2. *$(d_G(x_i, y_j, A))_{(i,j)\in\mathcal{S}\backslash s}$ are identically distributed, and their common cumulative distribution function is f_G.*
3. *$(d_D(D(x_i), D(y_j)))_{(i,j)\in\mathcal{S}}$ are identically distributed, and their common cumulative distribution function is f_D.*

Since the joint probability of independent variables is the product of the marginal probabilities, the probability defined in (2.4) can be calculated as follows.

Proposition 2.2.

$$p(\mathcal{S}, A, \delta_G, \delta_D) = f_D(\delta_D)^k f_G(\delta_G)^{k-m}, \tag{2.5}$$

where k is the cardinality of $\mathcal{S}$, and m is the cardinality of s.

In the hypothesis testing paradigm, hypothesis $\mathcal{H}_0$ is rejected as soon as $p(\mathcal{S}, A, \delta_G, \delta_D)$ is below the predetermined significance level. However, here it would mean that, all things being equal, large groups $\mathcal{S}$ would be favored since large k produces small probabilities in (2.5). Following the a contrario method, we do not deal directly with the probabilities but rather with the so-called *number of false alarms* (NFA), which permits us to discard the arbitrary significance level. The NFA corresponds to the average number of groups consistent with A, δ_G, δ_D under hypothesis $\mathcal{H}_0$. The NFA is estimated by multiplying the probability of a false alarm $p(S, A, \delta_G, \delta_D)$ by the number of possible events. Here there are $\min\{N_1, N_2\} - m$ choices for $k \geq m$, $\binom{N_1}{k}$ choices for the interest points in image 1, $\binom{N_2}{k}$ choices for the interest points in image 2, $k!$ choices for the correspondences, and $\binom{k}{m}$ choices for the minimal set to estimate A. Each minimal set s possibly leads to $Q = 3$ fundamental matrices (with the seven-point algorithm) or $Q = 1$ homography.

The following definition was outlined in [17] (*colored rigidity*) but was not studied further.

Definition 2.3. *We say that a set $\mathcal{S}$ of correspondences is ε-meaningful if there exist*

1. *two thresholds δ_G and δ_D such that*

$$\forall (i,j) \in \mathcal{S}, \quad d_G(x_i, y_j, A) \leq \delta_G, \tag{2.6}$$

$$\forall (i,j) \in \mathcal{S}, \quad d_D(D(x_i), D(y_j)) \leq \delta_D, \tag{2.7}$$

2. *a transformation A evaluated from m correspondences belonging to $\mathcal{S}$, such that*

$$NFA(\mathcal{S}, A, \delta_G, \delta_D) := Q\,(\min\{N_1, N_2\} - m)\,k! \binom{N_1}{k}\binom{N_2}{k}\binom{k}{m} f_D(\delta_D)^k f_G(\delta_G)^{k-m} \leq \varepsilon, \tag{2.8}$$

where k is the cardinality of $\mathcal{S}$, $m = 4$ and $Q = 1$ if A is a homography, and $m = 7$ and $Q = 3$ if A is a fundamental matrix.

Since f_D and f_G are nondecreasing, the following proposition holds.

Proposition 2.4. *A set $\mathcal{S}$ of correspondences is ε-meaningful if there exists a transformation A estimated from m correspondences belonging to $\mathcal{S}$ such that*

$$NFA(\mathcal{S}, A) := Q\,(\min\{N_1, N_2\} - m)\,k! \binom{N_1}{k}\binom{N_2}{k}\binom{k}{m} f_D(\delta_D)^k f_G(\delta_G)^{k-m} \leq \varepsilon, \tag{2.9}$$

where k is the cardinality of $\mathcal{S}$, $\delta_G = \max_{(i,j)\in\mathcal{S}} \max\{d_G(y_j, Fx_i), d_G(x_i, F^T y_j)\}$, and $\delta_D = \max_{(i,j)\in\mathcal{S}}(d_D(D(x_i), D(y_j)))$.

The aim of the algorithm discussed in section 3 is to find the most (or a very) meaningful set of correspondences, that is to say, the set of correspondences $\mathcal{S}$ with the lowest (or a very

low) NFA($\mathcal{S}$). Equation (2.9) balances the probability $f_D(\delta_D)^k f_G(\delta_G)^{k-m}$ and the number of possible sets of size k among the N_1 interest points from image 1 and the N_2 interest points from image 2. If δ_D and δ_G are fixed, when k grows, the former vanishes while the latter tends to increase (see Proposition A.1 in the appendix).

In the following sections we specify the choices for the distances d_D and d_G and for the associated cumulative distribution functions f_D and f_G.

2.2. Modeling the geometric constraint. In the $A = F$ case, Moisan and Stival [17] propose defining $d_G(y, Fx)$ as the Euclidean distance between y and the epipolar line Fx. They show that the probability for a random point y to lie at a distance less than δ_G from the line Fx is approximately $2D\delta_G/S$, where D and S are, respectively, the diameter and surface area of both images (assumed here to have the same size). In the present framework, with (2.2) and the independence assumption,

$$\Pr(d_G(x, y, F) \leq \delta_G) = \left(\frac{2D}{S}\delta_G\right)^2. \tag{2.10}$$

As we have seen earlier, $f_G(\delta_G)$ is balanced by the probability $f_D(\delta_D)$ related to the photometric constraint. We decide to parametrize the distribution f_G by using

$$f_G(\delta_G) = \left(\frac{2D}{S}\delta_G\right)^{2\alpha}. \tag{2.11}$$

We discuss the influence of this $\alpha > 0$ parameter in section 4.1.1.

Let us note that $\frac{2D}{S}\delta_G$ may be larger than 1 since it is actually an upper bound of the cumulative distribution function. In order to speed up the search, we decide to a priori eliminate groups such that this probability is larger than 5%. For typical 500×500 images, this corresponds to $\delta_G > 12.5$ pixels.

In the $A = H$ case, we just have to adapt the definition of $f_G(\delta_G)$ from a point-line correspondence (2.11) to a point-point correspondence as in [3, 23]:

$$f_G(\delta_G) = \left(\frac{\pi\delta_G^2}{S}\right)^{2\alpha}. \tag{2.12}$$

Indeed, $\pi\delta_G^2/S$ is the probability for a random point uniformly distributed across an image (surface area S) to be located at a distance less than δ_G from a fixed point.

2.3. Modeling the photometric constraint. Now we define d_D and f_D. Since the space of descriptors is neither isotropic nor homogeneous, any "good" metric over descriptors should take into account the vicinity of $D(x)$. Taking our inspiration from [22], and based on previous works [20], we define

$$d_D(D(x), D(y)) = \phi_{D(x)}(\mathsf{dist}(D(x), D(y))), \tag{2.13}$$

where dist is a distance over the descriptor space, and $\phi_{D(x)}$ is the cumulative distribution function of $\mathsf{dist}(D(x), D(\cdot))$ when $D(\cdot)$ spans the set of descriptors in image $\mathcal{I}_2$.

Note that, provided that $\phi_{D(x)}$ is continuous and increasing and $\mathsf{dist}(D(x), D(y))$ is a realization of the underlying random process, $d_D(D(x), D(y))$ is uniformly distributed over the unit interval $[0, 1]$ (this is a classic property of cumulative distribution functions; see [25]). This distance therefore automatically adapts to the heterogeneity of the descriptor space as a contextual dissimilarity measure. Moreover, we can simply set $f_D(t) = t$.

SIFT descriptors are made of $N = 16$ histograms, each with $m = 8$ bins discretizing the gradient orientation over the circular interval $[0, 2\pi)$. Rabin, Delon, and Gousseau [22] thus define the distance between descriptors as

$$\mathsf{dist}(D(x), D(y)) = \sum_{i=1}^{N} \mathsf{CEMD}(D^i(x), D^i(y)), \tag{2.14}$$

where CEMD is the circular earth mover's distance, which is well adapted for circular histograms [22].

For every $i \in [1, N]$, let us denote by $\varphi_{D^i(x)}$ the distribution function of $\mathsf{CEMD}(D^i(x), D^i(\cdot))$. Following the discussion in [22] and under the independence assumption, $\phi_{D(x)}$ is defined as

$$\phi_{D(x)}(\delta) = \int_0^{\delta} \bigotimes_{i=1}^{N} \varphi_{D^i(x)}(t)dt, \tag{2.15}$$

where $\otimes$ is the convolution product. Indeed, $\mathsf{CEMD}(D(x), D(y))$ appears as the sum of N random variables whose probability distribution is the convolution product of the N marginal distributions under the independence assumption. In practice, the distribution function $\varphi_{D^i(x)}$ is empirically estimated over the set of all $D^i(y)$ when y spans the set of the interest point extracted from image $\mathcal{I}_2$.

2.4. Discussing the NFA criterion. Sets with small NFA are the most relevant ones, as soon as the NFA is below 1. In this section, the discussion completes the comments on the so-called *colored rigidity* in [17]. For the sake of simplicity, we consider here the epipolar case and assume that $N_1 = N_2 = N$.

Let us note that

$$M(k, N) := 3(N - 7)k!\binom{N}{k}^2\binom{k}{7}. \tag{2.16}$$

Figure 1 shows the graph of $-\log_{10}(M(k, N))/k$ versus k for several values of N. From (2.9), this gives the maximal value for the logarithm of $f_D(\delta_D)f_G(\delta_G)^{1-7/k} \simeq f_D(\delta_D)f_G(\delta_G)$ so that a group $\mathcal{S}$ is 1-meaningful (in the case $N_1 = N_2 = N$). Indeed,

$$NFA(\mathcal{S}) \leq 1 \quad \text{iff} \quad \log_{10}\left(f_D(\delta_D)f_G(\delta_G)^{1-7/k}\right) \leq -\log_{10}(M(k, N))/k. \tag{2.17}$$

The NFA criterion meets two natural requirements:

- When N is fixed, the smaller the k, the smaller the probability product should be. This situation can be met when dealing with a large rate of outliers and seeking meaningful groups with small k with respect to N. Since f_D and f_G are nondecreasing, this means that thresholds δ_D and δ_G must be tighter in this case.

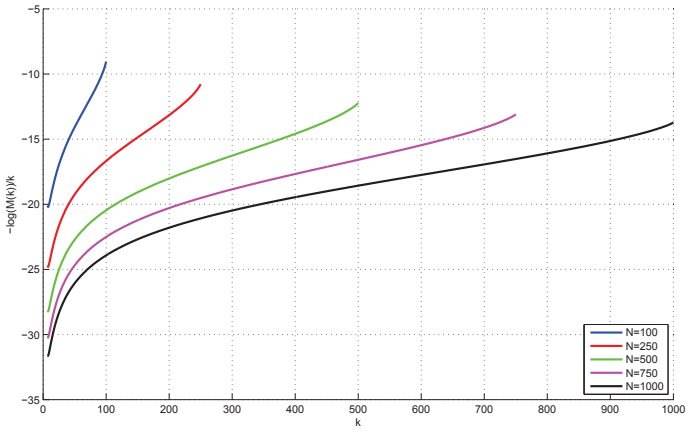

Figure 1. *$-\log_{10}(M(k,N))/k$ versus k for several values of N. The figure gives an approximation of the logarithm of $f_D(\delta_D)f_G(\delta_G)$ so that it is still possible to find a 1-meaningful set of correspondences.*

- When k/N is fixed, the larger the N, the smaller the probability product (and hence the tighter the thresholds) should be. This is handy when looking for correspondences in fixed size images: the denser the correspondences are distributed, the tighter δ_D and δ_G are.

3. A random sampling algorithm for finding meaningful sets. In this section we design a random sampling algorithm to obtain meaningful groups of correspondences, based on heuristic criteria. In section 3.1, an a contrario model (from [22]) is used to restrict the set of potential correspondences for a given interest point from image $\mathcal{I}_1$. Since the set of possible correspondences is still huge, a random sampling method is described in section 3.2.

3.1. Combinatorial reduction. In order to reduce the computational burden, we do not consider all possible correspondences $y_1, \dots, y_{N_2}$ in image $\mathcal{I}_2$ for an interest point x_i from image $\mathcal{I}_1$, but only the set of tentative correspondences $y_{j_1}, \dots, y_{j_{N_i}}$ such that the distance between the associated descriptors is below some threshold. This threshold should be large enough so that the true matching decision is not made at this step, while eliminating clearly spurious correspondences. We use the a contrario framework from [22]. In this case y_j is a tentative correspondence to x_i if, with the notation of (2.13),

$$N_1 N_2 d_D(D(x_i), D(y_j)) \leq \widetilde{\varepsilon}. \tag{3.1}$$

The value of $\widetilde{\varepsilon}$ does not depend on the experimental setup and is carefully discussed in [22]. In this paper, we set $\widetilde{\varepsilon} = 10^{-2}$, which gives a reasonable number of tentative correspondences $y_1, \dots, y_{j_N}$. This choice is motivated in section 4.1.2. In practice, we get between 0 and 30 tentative correspondences for each x_i in a typical image.

3.2. Random sampling algorithm. The aim is to select one (or zero) $y_{j(i)}$. We use a random sampling algorithm. It is a two-step iterative algorithm, which we describe for the two cases of interest (fundamental matrix F or homography H) as follows:

Step A. Draw a sample made of seven correspondences for estimating F, or four for H.

Step B. Look for the most meaningful group made from a subset of the preceding tentative correspondences, consistent with F or H.

Step A. Drawing a seven- or four-correspondence sample. Seven (or four) points x_i are uniformly drawn and then are associated with a tentatively corresponding point $y_{j(i)}$. Since it gives good experimental results and reduces the computational burden, we use descriptor nearest neighbor matching.

Up to three fundamental matrices are estimated via the nonlinear "seven-point algorithm" [11]. In the case of the homography, the transformation is estimated by the direct linear transform solved by singular value decomposition.

Note that the SIFT algorithm may extract several keypoints at the same location but with different orientations or scales. In order to avoid degenerate cases, we check that the minimum sample does not contain such points.

Step B. Seeking meaningful groups. Correspondences are added to the previous seven (or four) correspondences to form as meaningful a group as possible. To that end, we use of the following heuristic, which consists in iterating the following stages:

1. For every x_i, select

$$y_{j(i)} = \underset{y_{j_k}}{\operatorname{argmin}} \{f_D(d_D(D(x_i), D(y_{j_k}))) \cdot f_G(d_G(A, x_i, y_{j_k}))\} \tag{3.2}$$

   and sort correspondences $(x_i, y_{j(i)})$ in increasing order of this product of probabilities, in order to obtain a series of nested groups made of $k = 7, 8, 9, \ldots, N_1$ correspondences. This step can produce correspondences between $N > 1$ x_i's and a single y_j, which is unrealistic in two-view matching. Therefore, we decide to keep among these correspondences a single one, namely $(x_i, y_{j(i)})$, such that the above-mentioned probability product is minimized.
2. Compute the NFA for each of the above-mentioned nested groups and select the most meaningful one.
3. Sort correspondences $(x_i, y_{j(i)})$ in increasing order of $f_G(\delta_G(A, x_i, y_{j(i)}))$ to build a new set of nested groups, compute the NFA, and select the most meaningful one.
4. Return the most meaningful group found in either step 2 or 3.

Steps 1 and 2 obviously do not ensure that the obtained group is the most meaningful one with a fixed F matrix (unlike the a contrario RANSAC algorithm in [17] where only the geometric criterion is used). This heuristic is based on the fact that, provided k is fixed, the most meaningful group minimizes the product $f_D(\delta_D) f_G(\delta_G)$. Note that step 1 permits us to select correspondences among nonnearest neighbors. We have experimentally remarked that step 3 often allows us to discard false correspondences that are introduced with a low k in step 1 because the photometric distance is very short and overwhelms the (large) geometric distance.

In Step A, we could also have avoided biasing the algorithm by the nearest neighbor choice. As in [29], it would be possible to pick up for each i the corresponding point $y_{j(i)}$ by drawing it randomly in the set $y_{j_1}, \ldots, y_{j_{N_i}}$, where y_{j_i} has weight $K/d_D(D(x_i), D(y_{j_i}))$ (K is a normalization parameter). This scheme would preferably select nearest neighbors but also permit nonnearest neighbors among the minimum sample. However, the outlier

Input: two views of the same 3D scene.

1. Use SIFT algorithm to extract interest points and invariant descriptors from each view: $(x_i, D(x_i))_{i\in\{1,\dots N_1\}}$ and $(y_j, D(y_j))_{j\in\{1,\dots N_2\}}$.
2. For every $i \in \{1, \dots N_1\}$,
   (a) build the empirical distance d_D (as explained in section 2.3),
   (b) define a set of tentative correspondences (section 3.1).
3. Iterate ($N = 20,000$):
   (a) Choose seven (*resp., four*) points x_i and detect the seven *(resp., four)* corresponding points $y_{j(i)}$ (heuristic A in section 3.2).
   (b) Compute the three possible fundamental matrices F from these seven correspondences and go to (c) for each of these matrices *(resp., compute the homography H from these four correspondences and go to (c))*.
   (c) Select the most meaningful group (heuristic B in section 3.2).

Output: the most meaningful set of correspondences ever encountered.

Figure 2. *The proposed a contrario matching algorithm.*

rate is significantly larger for nonnearest neighbors (which can be verified in experiments; see Table 3); this last scheme thus needs many more iterations without improving the results.

The number N of iterations of this sampling strategy is set at a quite large value ($N = 20,000$), although it could be tuned based on a prior estimation of the outlier rate.

To sum up the discussion, the whole algorithm is described in Figure 2.

4. Experiments.

The algorithms are tested on pairs of images of the same scene, seen from two different viewpoints. These images are either natural or computer generated. They have repeated patterns in order to illustrate the interest of detecting correspondences beyond the first nearest neighbor. Since SIFT features are not robust to wide viewpoint changes, we consider only limited viewpoint changes. The application of the a contrario model to wide viewpoint changes is discussed in [21]. The assessment is based on visual inspection of the retrieved correspondences as in [19] and display of the epipolar lines. The correspondences are depicted either by a line between the matching points in both images or by a straight segment whose length is the apparent motion of the point in a view to the corresponding point in the other view. Note that comparing to a ground truth, such as the camera position estimated from the fundamental matrix, is not an easy task and requires some Euclidean information on the scene, which is beyond the scope of this paper.

The experiments are organized in the following way. Section 4.1 discusses the two modeling parameters of the a contrario matching, namely α (section 2.2) and $\widetilde{\varepsilon}$ (section 3.1). Section 4.2 shows that the a contrario matching algorithm retrieves correspondences even with a large number of repeated patterns. It is also compared with an implementation of generalized RANSAC. These experiments show that the matching thresholds (δ_D and δ_G) are automatically derived and actually vary, and that we are able to select correspondences that are not nearest neighbors. Although generalized RANSAC with the correct parameters gives similar results in most experiments, these parameters must be tuned for each experiment

after a visual assessment of the result. The a contrario matching, for its part, is based on the NFA which directly quantifies the relevance of each set of correspondences.

The algorithmic complexity of each iteration is $\mathcal{O}(CN_1)$, where $C \leq N_2$ is the maximum number of tentative correspondences for each point in the first view. Building the empirical distance d_D has a complexity of $\mathcal{O}(N_1 N_2)$, and the algorithm of [22] to build the set of tentative correspondences has the same complexity. The computation time of the a contrario algorithm is about 30–40 seconds for typical 500×500 images ($\simeq$ 5 seconds for 200×200 images). Speeding up would be possible via multicore programming and by enhancing the estimation of the distribution function of the distance between descriptors (the $\varphi_{D^i(x)}$ in (2.15)), e.g., by subsampling the dataset in a Monte Carlo estimation.

4.1. Sensitivity of the a contrario model to the parameters. Here we test the influence of the parameters α (from section 2.2) and $\widetilde{\varepsilon}$ (from section 3.1).

4.1.1. Influence of α. From (2.9), the parameter α in f_G ((2.11) or (2.12)) permits us to balance between the geometric and photometric probabilities. These probabilities do not have the same order of magnitude: the former varies around 10^{-5}, while the latter may be around 10^{-20}. Thus α behaves as a normalization parameter.

Equations (2.11) and (2.12) show that the smaller α is, the smaller the contribution of the geometric constraint in the NFA is. As a matter of fact, a smaller α generates a set of correspondences that is less constrained by the geometry and more constrained by the photometric resemblance of the descriptors. Conversely, a larger α should generate groups of correspondences that meet the geometric constraint well; however, the photometric constraint may be too large in this case, and the geometry may not correspond to the reality. Moreover, the geometric and photometric constraints are offset by the number of corresponding points as explained in section 2.4: a small group can outweigh a large one if the photometric (small α) or geometric (large α) constraint is tightly enforced. A trade-off appears, as illustrated in Figures 3 and 4.

Table 1 provides statistics for when α grows and thus the geometric constraint has more importance. One can see that when α grows from 1 to 5, the number of retrieved correspondences grows, and the geometric accuracy is better (here, distance to the epipolar line), while the photometric constraint becomes less tight. Note that α larger than 6 does not produce significant changes in the most meaningful set of correspondences; basically the same points are actually retrieved. The same experiment with the homography constraint offers similar results (not shown here). From this table and other experiments, we decide to set $\alpha = 5$ in all following experiments. With such a value, the probability product $f_D(\delta_D) f_G(\delta_G)$ can attain the magnitude required in Figure 1.

In these experiments, $\widetilde{\varepsilon}$ was set to 10^{-2}.

4.1.2. Influence of $\widetilde{\varepsilon}$. Here we test the influence of the $\widetilde{\varepsilon}$ parameter (section 3.1). Since SIFT descriptors are invariant to similarity only, a bias will appear in the probabilities in the case of large viewpoint changes, as in every SIFT-based method. We therefore consider a small motion between two views, so that the effect of the limited invariance of SIFT to viewpoint change is as low as possible. The test is performed here with the fundamental matrix model. As an illustration, Figure 5 shows results with a varying $\widetilde{\varepsilon}$.

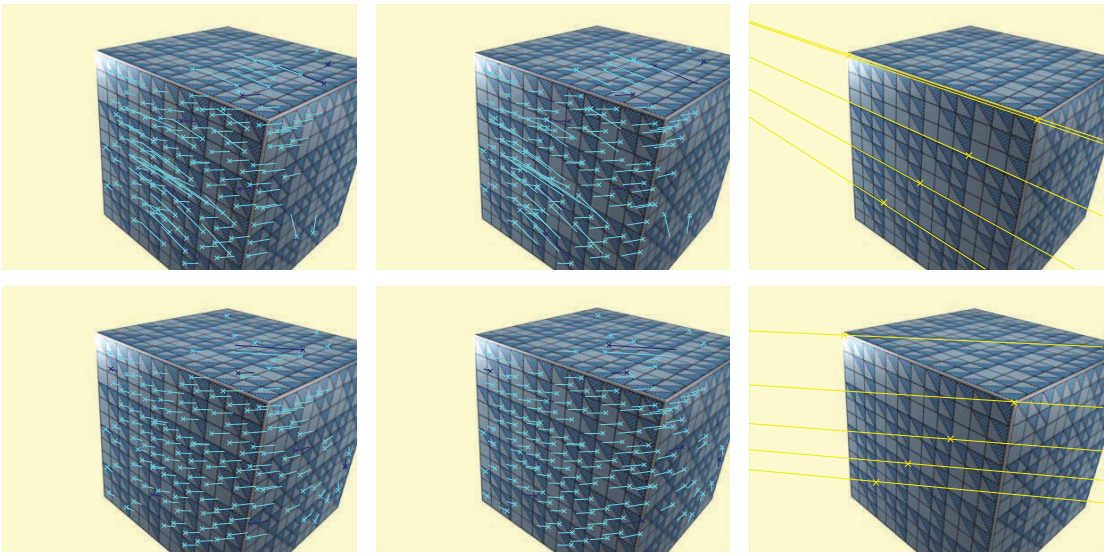

Figure 3. *Influence of the α parameter, epipolar constraint. Top: $\alpha = 3$. There are 138 correspondences (between the left and middle images; 31 of them are not nearest neighbors). In this case, the geometric constraint is not imposed in a strong enough way. The most meaningful group corresponds to the situation where many correspondences are found along the dominant direction of the lattice of repeated patterns (vanishing lines as shown in the right image; see the discussion in section 4.2.1). Bottom: $\alpha = 5$. There are 148 correspondences (47 are not nearest neighbors), and the retrieved epipolar pencil is now consistent with the camera motion. Note a false correspondence on the top of the cube that is consistent with respect to photometry (matches between repeated patterns) and geometry (matches along an epipolar line) as in Figure 6.*

Table 2 shows statistics from this experiment. One can see that reducing $\widetilde{\varepsilon}$ also reduces the number of tentative correspondences among which the most meaningful set is sought, while having almost no impact on the cardinality of this set. In other words, decreasing the value of $\widetilde{\varepsilon}$ speeds up the search and discards mainly false correspondences. Note that at least 20–25% of the matches are not nearest neighbors.

From these results and other experiments on realistic images, we set $\widetilde{\varepsilon} = 10^{-2}$.

4.2. Point correspondences and repeated patterns. Section 4.2.1 discusses the specific problems raised by repeated patterns. Section 4.2.2 describes an implementation of generalized RANSAC (inspired by [29]), and section 4.2.3 discusses the output of the a contrario matching algorithm.

4.2.1. Matching images with repeated patterns under epipolar constraint. We focus on images where repeated patterns are present, like windows on a facade or a repetitive texture. As discussed in the introduction, erroneous correspondences can be obtained because the descriptors are alike and the pair of matching points is kept by RANSAC as soon as they lie on the same epipolar plane. The psychophysics community has been aware of this situation, known as the *double nail illusion* after a paper by Krol and van de Grind [12], since the 1980s.

In some situations, the double nail illusion can even fool the matching algorithm and return spurious epipolar lines which are actually vanishing lines (see Figure 6.) Indeed, any

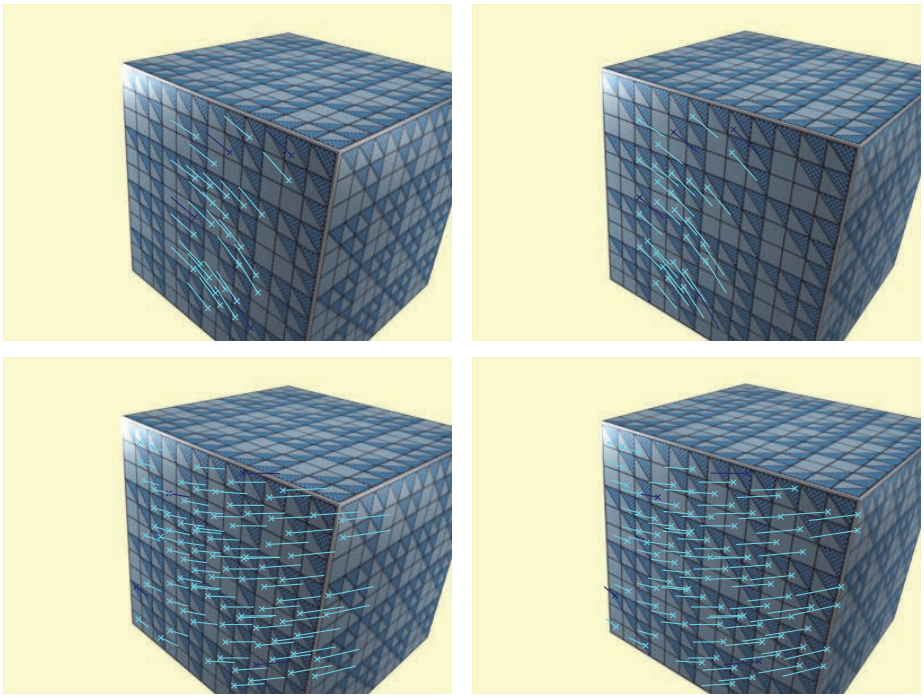

Figure 4. *Influence of the α parameter, homographic constraint. Top: $\alpha = 1$. There are 31 correspondences between the left and right images (four are not nearest neighbors). Corresponding features are actually perceptually similar but shifted. Such correspondences are possible here because of the large number of repeated patterns. Here the photometric constraint has a too strong influence on the NFA and is not balanced by the geometric constraint. Bottom: $\alpha = 3$. Ninety-one correspondences are retrieved (56 are not nearest neighbors): the influence of the geometry is strengthened and the algorithm provides a larger group, which is now correct.*

Table 1

Influence of α on the retrieved sets of correspondences, using the same images as in Figure 3. From left to right: α, No. of points (the cardinality of the retrieved set), δ_G, and $\log(\delta_D)$. Standard deviations are indicated between parentheses (averaged over 100 runs). Up to $\alpha = 6$, the larger α is, the smaller δ_G is. When α is between 6 and 10, the accuracy does not decrease anymore because there are only a few matches with a distance to the epipolar line less than 0.4. Groups with still lower NFA are then retrieved, but with the same k, δ_G, δ_D.

α	No. of points	Geometry δ_G in pixel	Photometry $\log(\delta_D)$
1	149.2 (1.8)	.71 (.12)	−40.4 (.5)
2	151.6 (3.8)	.70 (.14)	−39.5 (1.1)
3	162.1 (8.1)	.58 (.12)	−33.6 (3.2)
4	164.2 (5.2)	.49 (.13)	−30.5 (1.8)
5	172.3 (10)	.46 (.11)	−26 (4.2)
6	183.5 (6.2)	.49 (.09)	−21.2 (1.9)
7	182.9 (5.3)	.47 (.10)	−20.7 (.48)
8	183.7 (4.9)	.46 (.08)	−20.6 (.28)
9	184.5 (4.5)	.46 (.08)	−20.5 (.28)
10	185 (4.8)	.46 (.08)	−20.5 (.17)

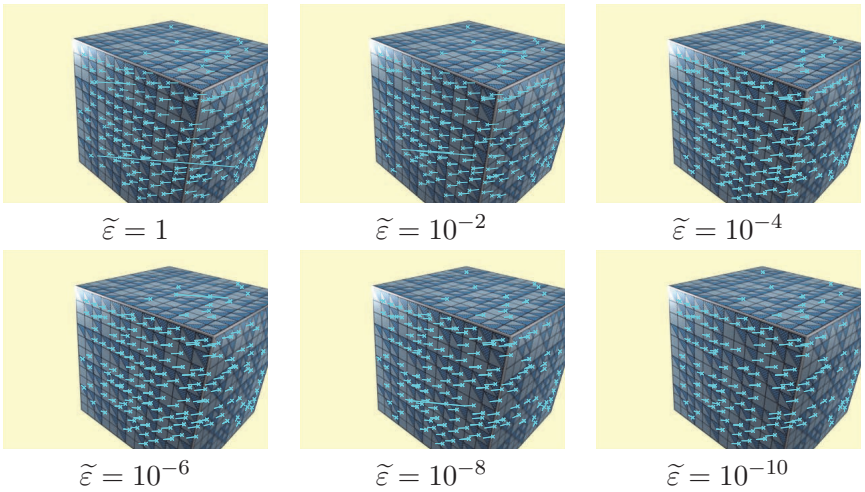

Figure 5. Synthetic *images. The most meaningful group consistent with a fundamental matrix is sought for six values of* $\widetilde{\varepsilon}$. *We show only the first view. The blue segment corresponds to the apparent motion of an interest point (localized by a cross) between the two views. Some false correspondences are still retrieved. A careful examination shows that they actually lie along the associated epipolar line. In all experiments (whatever the distance and* $\widetilde{\varepsilon}$ *as in Table 2), the average distance to the epipolar line is about* 0.2–0.3 *pixel.*

Table 2

Influence of $\widetilde{\varepsilon}$. *From left to right: the six values of* $\widetilde{\varepsilon}$ *in the range* 1–10^{-10}, *the number of tentative correspondences retrieved after the combinatorial reduction step (section 3.1), the cardinality of the most meaningful group (averaged over* 100 *runs, with standard deviation in brackets), and the average proportion of nearest neighbors.*

$\widetilde{\varepsilon}$	No. tentative corr.	No. most meaning. group	% of rank 1 corr.
1	2027	219.5 (6.4)	76.1
10^{-2}	1409	220.6 (4.9)	76.2
10^{-4}	999	218.3 (3.9)	76.6
10^{-6}	663	202.7 (3.0)	78.4
10^{-8}	407	172.8 (2.8)	85.5
10^{-10}	274	146.2 (2.2)	90.3

set of points lying on parallel 3D lines satisfies a degenerate epipolar constraint in which the epipoles are the vanishing points, whichever way the corresponding points are matched on the epipolar line (disregarding the 3D preimage). If the points on the epipolar lines have similar descriptors, then the degenerate situation can give the largest consensus set in RANSAC and wins over smaller but correct sets of correspondences.

4.2.2. A generalized RANSAC algorithm [29]. Standard RANSAC takes as input the tentative one-to-one point correspondences given by a preliminary step based on descriptor similarity. If this step is relaxed so that each interest point x in the first image has K tentative correspondences $y_1, \ldots, y_K$ in the second, then generalized RANSAC consists in iterating the following operations (here $y(x)$ is one of the $y_1, \ldots, y_K$ associated to y_i):

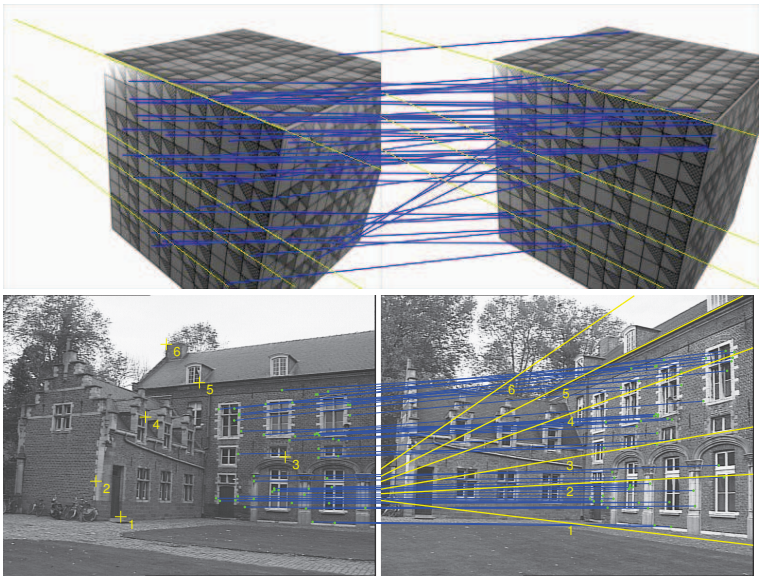

Figure 6. *The correspondences here satisfy both the descriptor similarity and the epipolar constraint. However, many false correspondences can be seen (in the lower pair, correspondences can be seen with the fourth windows, which is not visible in the first image), and the epipolar lines (in yellow) correspond to vanishing lines, giving an incorrect camera pose estimation.*

1. Draw a minimum sample $(x, y(x))$ to estimate H (or F).
2. Knowing H (or F), associate each x to a single $y(x)$ among the tentative correspondences.
3. Count the correspondences $(x, y(x))$ such that $d(x, y(x), H)$ (resp., $d(x, y(x), F)$) as in (2.2) (resp., (2.3)) is less than a predetermined threshold δ_{GR}.

In the end, the largest set of correspondences is returned.

Note that two parameters are involved: K and δ_{GR}. In step 1 we decide to associate x to its nearest neighbor $y(x)$ among the tentative correspondences y_i (in the sense of the descriptor proximity), and in step 2 to define $y(x)$ among the y_i as minimizing the distance d. In [29] it is suggested to randomly sample the x and $y(x)$, and different strategies are proposed. We found that taking the nearest neighbor in step 1 significantly reduces the number of needed iterations. This is a sound hypothesis since in tractable cases a large number of correct correspondences can be found among nearest neighbors (see Table 3).

Let us discuss a property of generalized RANSAC which was not mentioned in [29]. As discussed in the introduction, in some examples nearest neighbors conditioned by the distance ratio produce erroneous sets of correspondences. In Figure 7, we show that when relaxing the tentative correspondences to the nearest neighbors (without any condition on the distance ratio) it is possible to find a correct set of correspondences. Of course, a large number of iterations is needed since the outlier rate significantly increases. Here about 31,000 iterations are needed to obtain a correct set (averaged over 10 runs). However, when searching among $K = 5$ first nearest neighbors, only about 13,000 iterations are needed. For example, the group shown in Figure 7 (bottom) comprises 43 correspondences of rank 1, 25 of rank 2, 24 of

Table 3

Number of occurrences of the nth nearest neighbors selected by the AC method (homography case). Monkey *corresponds to Figure 9 (*412 *versus* 445 *extracted keypoints),* Loria *to Figure 10 (*2,562 *versus* 2,686*),* Flatiron *to Figure 11 (*756 *versus* 598*). The NN-T+O method does not succeed at all in the* Loria *(Figure 10) and* Flatiron *(Figure 11) experiments. Note the variability of the matching ranks. Ranks higher than* 2 *are more frequent when the scene contains repeated patterns and cannot be retrieved by the NN-T+O method or any method limited to nearest neighbor matching.*

	Number of correspondences		
Rank	*Monkey*	*Loria*	*Flatiron*
1	42	98	8
2	23	32	3
3	17	29	1
4	11	18	1
5	8	20	0
6	8	19	0
7	4	11	0
8	8	5	0
9	3	13	0
≥ 10	4	67	0
Total	128	312	13

rank 3, 14 of rank 4, and 7 of rank 5. The subset of correspondences of rank 1 is beaten by the erroneous set shown in Figure 7 (top), which comprises 49 matches. Thus, fewer iterations are needed when searching beyond the nearest neighbor. When searching among $K = 10$ first nearest neighbors (not shown), about $6,000$ iterations are needed to obtain a correct set comprising fewer than 50% of nearest neighbors. In this case there are only 4 correspondences between the 7th and 10th nearest neighbors. Despite each pattern being repeated more than 10 times, this means that the "invariant" descriptors are actually not repeated more than 5–6 times within the distance threshold range.

Figure 8 shows the correspondences with the present a contrario matching, where the distance thresholds and the number of tentative correspondences are automatically set.

4.2.3. Assessing the a contrario and generalized RANSAC algorithms. The a contrario algorithm is compared with the method mentioned in section 1 (step 2), i.e., NN-T matching (nearest neighbor matching based on the Euclidean distance, and threshold on the ratio set to 0.6 as in Lowe's code), followed by the a contrario RANSAC from [17], called ORSA. This two-step scheme is abbreviated as NN-T+O, and our method is abbreviated as AC for a contrario. We also compare with generalized RANSAC.

Repeated patterns and homography. We first investigate the homography case. The number of matches selected with NN-T is small, as shown in the right image of Figure 9: repeated features are generally discarded at this early stage, and of course cannot be retrieved by the subsequent RANSAC. As shown in the left image of Figure 9, the AC method retrieves many more correspondences. The numerous extra correspondences coincide with nonnearest neighbors (with respect to descriptor distance). From Table 3, while 128 features are matched, 42 are ranked first, and 86 have higher ranks.

We recall that the AC algorithm automatically adapts the thresholds to the scene: in the *Monkey* image pair (1,500 $\times$ 1,200 images, Figure 9) δ_G was derived as 3.1 pixels, in *Loria*

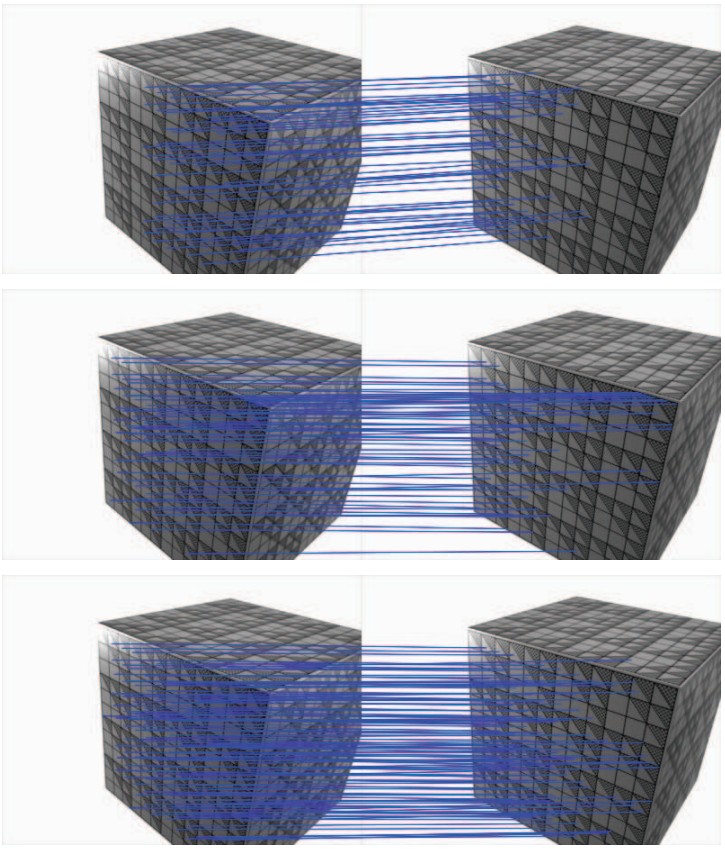

Figure 7. *Generalized RANSAC. About* 800 *SIFT keypoints are extracted from both images. The distance threshold* δ_{GR} *is set to* 1. *Top and middle:* $K = 1$ *and homography constraint. A large yet erroneous consensus set can be retrieved, here* 49 *matches (top). However, ensuring a large enough number of iterations in generalized RANSAC makes it possible to find correct sets, here* 60 *matches (middle). Bottom:* $K = 5$. *Here,* 113 *correspondences are retrieved, but only* 43 *are ranked first; fewer iterations are needed.*

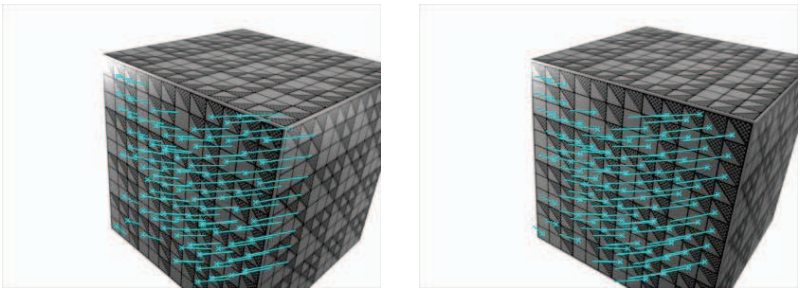

Figure 8. *Proposed a contrario matching; apparent motion of the interest points between the two views. All correspondences are correct. Here* δ_G *was determined as* 2.6 *pixels. Among the* 111 *matches,* 67 *are nearest neighbors,* 20 *are second nearest neighbors,* 10 *are third nearest neighbors, and the rest are fourth to eleventh nearest neighbors.*

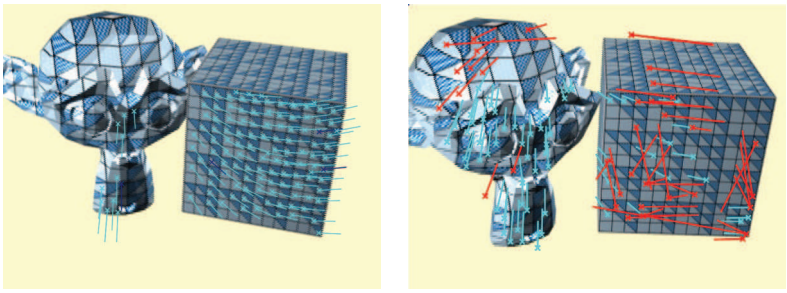

Figure 9. Monkey, *homographic constraint. Two images with repeated patterns. On the left, the proposed AC model. Most of the patterns lying on the dominant plane are detected. On the right, the second image with correspondences from NN-T (both colors) and NN-T+O (inliers in blue, outliers in red). Many more correspondences are retrieved with the AC algorithm. Generalized RANSAC gives similar results.*

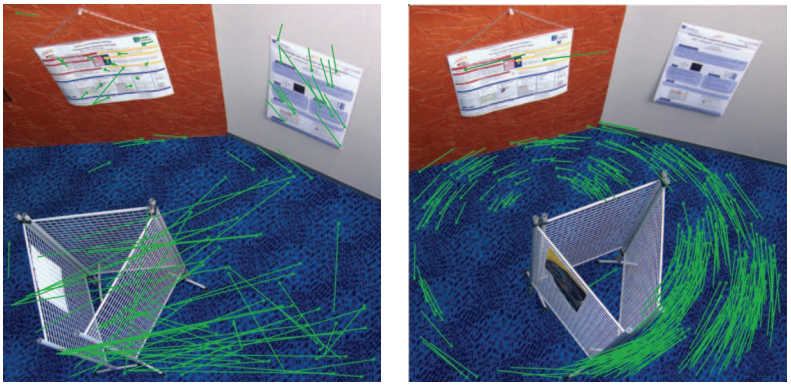

Figure 10. Loria *image pair. On the left, NN-T fails at identifying correct correspondences on the carpet. Thus, no subsequent RANSAC will succeed in drawing out the true correspondences. On the right, AC finds correct correspondences despite the large number of repeated patterns.*

(800 × 800 images, Figure 10) as 2.1 pixels, and in *Flatiron* (500 × 400 images, Figure 11) as 7.8 pixels. Using these values as distance threshold δ_{GR} in generalized RANSAC gives similar results, but this parameter (and K as well; see Table 3) has to be guessed.

Repeated patterns and epipolar constraint. We test the behavior of the AC algorithm under epipolar constraint. Figure 12 shows a situation with a repetitive texture and almost no false correspondence. Small baseline matching gives good results thanks to the limited invariance of the descriptors. In this case, the nearest neighbor is more likely to be the correct correspondence (as proved by the large proportion of matches among nearest neighbors) than in the case of a larger baseline. However, as soon as the baseline grows, generalized RANSAC, as well as AC matching, is trapped by the double nail illusion, and epipolar lines often degenerate into vanishing lines. All methods dealing only with points and invariant descriptors are prone to fail in this case, as discussed in section 4.2.1.

Figure 13 shows results of both the NN-T+O method (29 correspondences) and the AC method (92 correspondences). The AC matches are distributed in a denser fashion. The

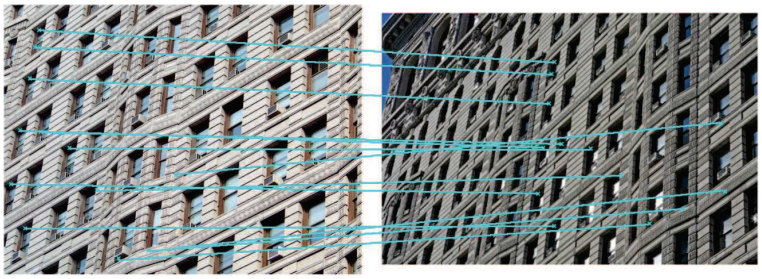

Figure 11. Flatiron, *homographic constraint. Thirteen matches can be found with the AC method.*

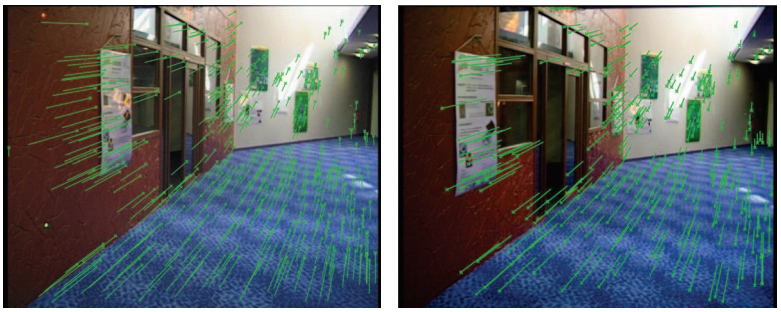

Figure 12. Corridor, *epipolar constraint. The AC method (left) retrieves* 423 *correspondences. Of these,* 405 *correspondences have rank* 1, 13 *rank* 2, 2 *rank* 3, 1 *rank* 4, 1 *rank* 6, *and* 1 *rank* 10. *The NN-T+O method (right) retrieves* 295 *out of* 316 *NN-T matches. The additional correspondences are on the carpet and on the wall.* 1,269 *keypoints were extracted from image* 1 *and* 1,360 *from image* 2.

repeated left-hand windows are not correctly matched with NN-T+O (marked in red in Figure 13), but they are all retrieved and not shifted with the AC method. However, several false correspondences are also retrieved, such as the ones between the structures of the left-hand facade which are indeed shifted along the epipolar lines (compare with the position of the windows; the same phenomenon appears on the right-hand facade). To assess the validity of the geometry, the fundamental matrix is re-estimated over the consensus set by minimizing the Sampson metric [11], and the hand-picked correspondences (in yellow) are compared with the associated epipolar lines. With the NN-T+O method, the distance is between 5 and 20 pixels. With the AC method, the distance is less than 5 pixels, except from one point on the parallelepipedic structure in the foreground which is still 15 pixels distant. The accuracy on this structure is poorer than the accuracy on the rest of the scene since a very small number of points is extracted from it, and since its apparent motion is quite different from the background's motion.

In some cases, motion estimation is inconsistent with the ground truth because of the repeated patterns, as explained in section 4.2.1. For example, in Figure 14 we can see that the most meaningful group consists in wrong correspondences among points that match in a dominant plane along lines parallel to an edge of the cube. Note that the stricter point-to-point constraint of the homography case (compare Figure 14 with Figure 9) enables us to retrieve a consistent set.

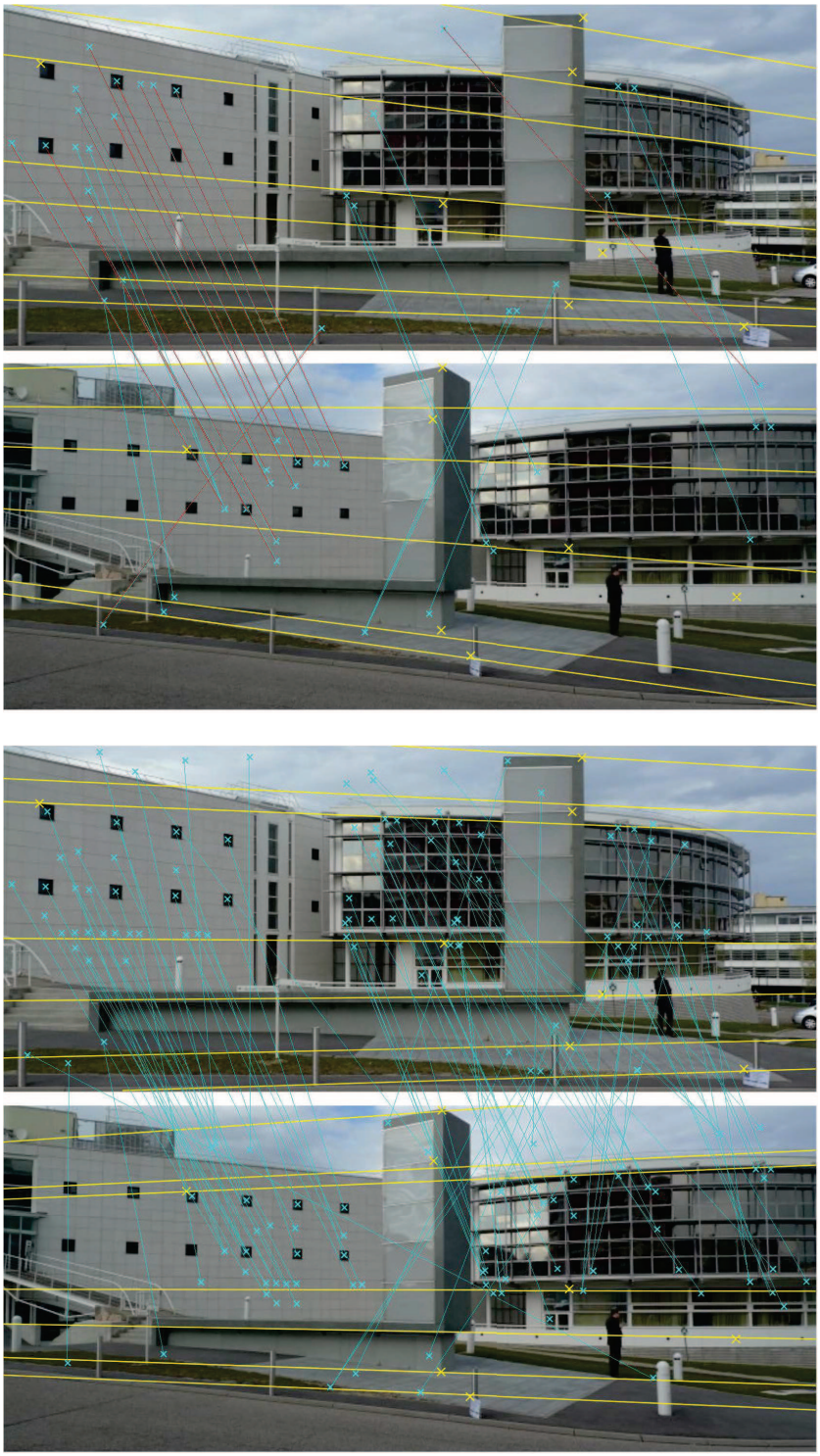

Figure 13. Loria building. *Top:* 29 *correspondences from the NN-T+O methods. Bottom:* 92 *correspondences from the AC method. The additional correspondences improve the reliability of the epipolar pencil. (See text for details).*

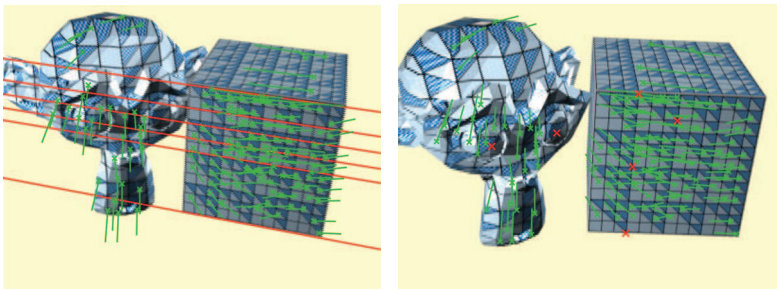

Figure 14. Monkey, *epipolar constraint. Failure case study. The recovered geometry corresponds to the vanishing lines. A few hand-inserted points in red show the epipolar lines pencil, which corresponds to the pattern alignment along the vanishing lines and not to the true motion. See section* 4.2.1.

5. Conclusion.

This paper extended past a contrario models for interest point matching by integrating both geometric and photometric constraints into the matching process. The resulting algorithm is able to detect correspondences beyond the first nearest neighbor, and it has adaptive matching thresholds. Experiments suggest that the values of its modeling parameters are not critical. The standard nearest neighbor approach can lead to erroneous matches, for example, in the case of images with repeated patterns. In this situation, the proposed a contrario approach is able to extract a set of consistent correspondences and performs as well as generalized RANSAC, which requires tuning several parameters.

Appendix. $M(k,N)$ increases up to $k_0 = N - \sqrt{N} + \mathcal{O}(\sqrt{N})$.

Proposition A.1. *Let N be an integer, and let for every integer $7 \leq k \leq N$: $M(k,N) = 3(N-7)k!\binom{N}{k}^2\binom{k}{7}$. The series $(M(k,N))_k$ is increasing between $k = 7$ and $k = k_0$, and decreasing for $k \geq k_0$, where*

$$k_0 = \left(2N + 1 - \sqrt{4N - 23}\right) \Big/ 2 \; =_{N\to+\infty} N - \sqrt{N} + \mathcal{O}(\sqrt{N}).$$

Proof. The ratio between two consecutive terms is $\frac{M(k+1,N)}{M(k,N)} = \frac{(N-k)^2}{k-6}$. This ratio is larger than 1 iff $P(k) = k^2 - (2N+1)k + N^2 + 6$ is positive, which is true, provided that $k < k_0$, where $k_0 = \left(2N + 1 - \sqrt{4N - 23}\right)/2$ is the smallest root of P. The second root is indeed larger than N, and $k \leq N$. ∎

REFERENCES

[1] H. Aanæs, A. L. Dahl, and K.S. Pedersen, *Interesting interest points: A comparative study of interest point performance on a unique data set*, Int. J. Comput. Vision, 97 (2012), pp. 18–35.

[2] W. Aguilar, Y. Frauel, F. Escolano, M.E. Martinez-Perez, A. Espinosa-Romero, and M.A. Lozano, *A robust Graph Transformation Matching for non-rigid registration*, Image Vision Comput., 27 (2009), pp. 897–910.

[3] T. Buades, Y. Lou, J.-M. Morel, and Z. Tang, *A note on multi-image denoising*, in Proceedings of the International Workshop on Local and Non-Local Approximation in Image Processing, Tuusalu, Finland, 2009.

[4] F. Cao, J.-L. Lisani, J.-M. Morel, P. Musé, and F. Sur, *A Theory of Shape Identification*, Lecture Notes in Math. 1948, Springer-Verlag, Berlin, 2008.

[5] J. Cech, J. Matas, and M. Perdoch, *Efficient sequential correspondence selection by cosegmentation*, IEEE Trans. Pattern Anal. Mach. Intell., 32 (2010), pp. 1568–1581.
[6] H. Deng, E. N. Mortensen, L. Shapiro, and T. G. Dietterich, *Reinforcement matching using region context*, in Proceedings of the Beyond Patches Workshop, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, 2006.
[7] A. Desolneux, *A probabilistic grouping principle to go from pixels to visual structures*, in Discrete Geometry for Computer Imagery, Lecture Notes in Comput. Sci. 6607, Springer, Berlin, Heidelberg, 2011, pp. 1–12.
[8] A. Desolneux, L. Moisan, and J.-M. Morel, *Meaningful alignments*, Int. J. Comput. Vision, 40 (2000), pp. 7–23.
[9] A. Desolneux, L. Moisan, and J.-M. Morel, *From Gestalt Theory to Image Analysis. A Probabilistic Approach*, Interdiscip. Appl. Math. 34, Springer, New York, 2008.
[10] M. Fischler and R. Bolles, *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography*, Comm. ACM, 24 (1981), pp. 381–395.
[11] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK, 2000.
[12] J.D. Krol and W.A. van de Grind, *The double-nail illusion: Experiments on binocular vision with nails, needles, and pins*, Perception, 9 (1980), pp. 651–669.
[13] C. Le Brese, J.J. Zou, and B. Uy, *An improved ASIFT algorithm for matching repeated patterns*, in Proceedings of the IEEE International Conference on Image Processing (ICIP), Hong Kong, 2010, pp. 2949–2952.
[14] D. Lowe, *Distinctive image features from scale-invariant keypoints*, Int. J. Comput. Vision, 60 (2004), pp. 91–110.
[15] K. Mikolajczyk and C. Schmid, *Scale & affine invariant interest point detectors*, Int. J. Comput. Vision, 60 (2004), pp. 63–86.
[16] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, *A comparison of affine region detectors*, Int. J. Comput. Vision, 65 (2006), pp. 43–72.
[17] L. Moisan and B. Stival, *A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix*, Int. J. Comput. Vision, 57 (2004), pp. 201–218.
[18] P. Moreels and P. Perona, *Evaluation of features detectors and descriptors based on 3D objects*, Int. J. Comput. Vision, 73 (2007), pp. 263–284.
[19] J.-M. Morel and G. Yu, *ASIFT: A new framework for fully affine invariant image comparison*, SIAM J. Imaging Sci., 2 (2009), pp. 438–469.
[20] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel, *An a contrario decision method for shape element recognition*, Int. J. Comput. Vision, 69 (2006), pp. 295–315.
[21] N. Noury, F. Sur, and M.-O. Berger, *How to overcome perceptual aliasing in ASIFT?*, in Advances in Visual Computing, Lecture Notes in Comput. Sci. 6453, Springer, Berlin, Heidelberg, 2010, pp. 231–242.
[22] J. Rabin, J. Delon, and Y. Gousseau, *A statistical approach to the matching of local features*, SIAM J. Imaging Sci., 2 (2009), pp. 931–958.
[23] J. Rabin, J. Delon, Y. Gousseau, and L. Moisan, *MAC-RANSAC: A robust algorithm for the recognition of multiple objects*, in Proceedings of the Fifth International Symposium on 3D Data Processing, Visualization and Transmission (3DPTV), Paris, France, 2010.
[24] E. Serradell, M. Özuysal, V. Lepetit, P. Fua, and F. Moreno-Noguer, *Combining geometric and appearance priors for robust homography estimation*, in Proceedings of the 11th European Conference on Computer Vision (ECCV), Part III, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 58–72.
[25] F. Sur, N. Noury, and M.-O. Berger, *Image Point Correspondences and Repeated Patterns*, Research Report 7693, INRIA, Nancy, France, 2011.
[26] B.J. Tordoff and D.W. Murray, *Guided-MLESAC: Faster image transform estimation by using matching priors*, IEEE Trans. Pattern Anal. Mach. Intell., 27 (2005), pp. 1523–1535.
[27] B. Triggs and P. Bendale, *Epipolar constraints for multiscale matching*, in Proceedings of the British Machine Vision Conference (BMVC), Aberystwyth, UK, BMVA Press, Manchester, UK, 2010.

[28] A. Vedaldi and S. Soatto, *Local features, all grown up*, in Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), New York, 2006, pp. 1753–1760.

[29] W. Zhang and J. Kosecka, *Generalized RANSAC framework for relaxed correspondence problems*, in Proceedings of the Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), Chapel Hill, NC, IEEE Computer Society, Washington, DC, 2006, pp. 854–860.